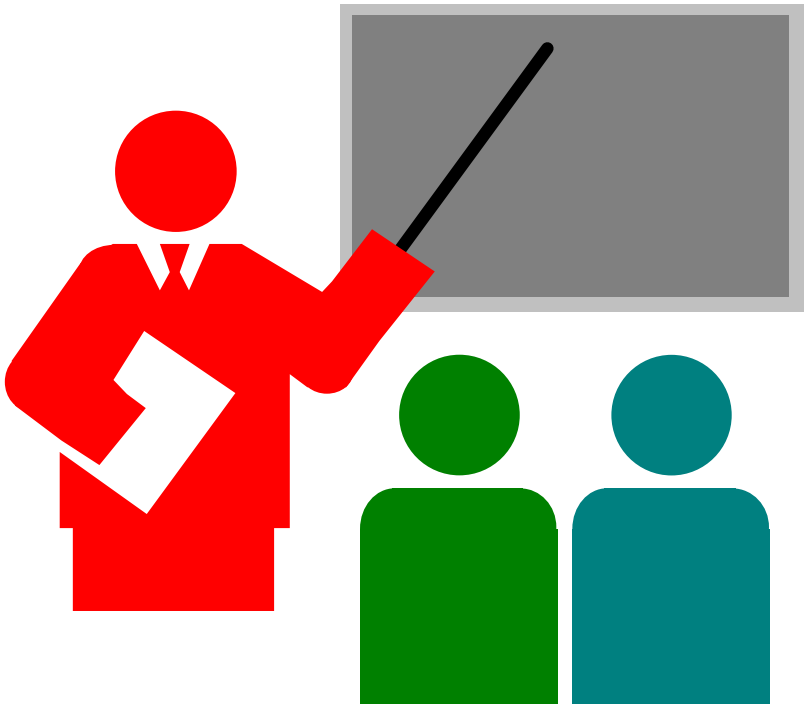
SOME CURRENT METHODOLOGICAL CHALLENGES IN MEDICAL PRODUCT EVALUATION

Larry Kessler, Sc.D.

Director, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, FDA

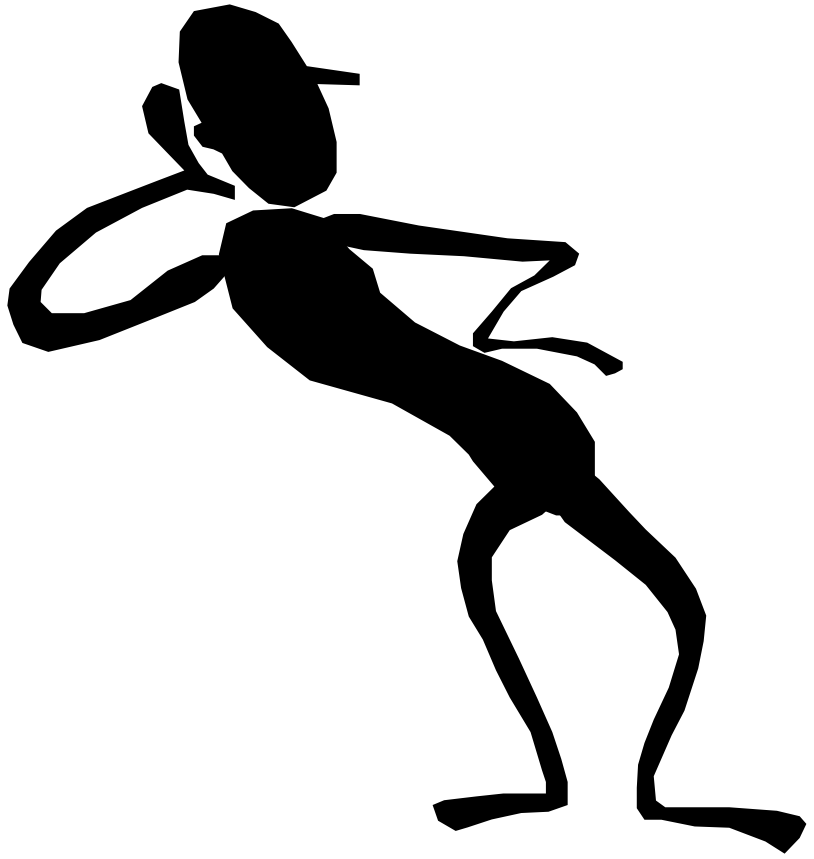
A Presentation in Three Parts



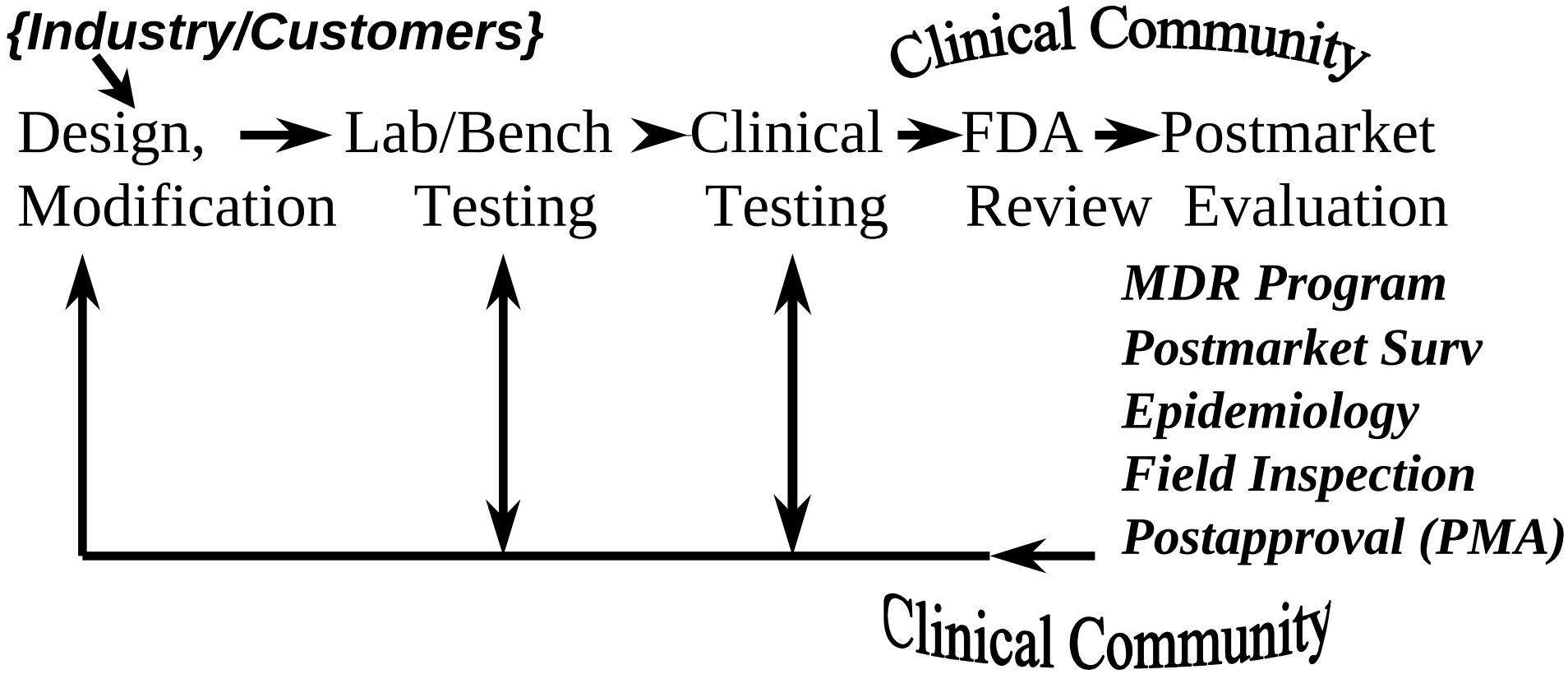
- An introduction to FDA from CDRH/OSB
- Drug and device trials: what constitutes sufficient evidence: methodologic challenges
- Postmarket paradoxes and more challenges

A Few Caveats

- This is my perspective
- I have worked for only four years as Director of the Office of Surveillance and Biometrics
- No secrets will emerge
- This presentation will be put on the FDA Website



From Design to Obsolescence: Medical Devices and Center for Devices and Radiological Health, FDA



'Design' ⟶ Device evolution ⟶ 'Obsolescence'

Case Studies

- How many trials?
- Bayesian methods and device trials
- Challenges and paradoxes in postmarket

Case Study 1: How Many Trials?

- Arguments have no chance against petrified training; they wear it as little as the waves wear a cliff
  - A Connecticut Yankee in King Arthur's Court by Mark Twain

The 1962 Kefauver-Harris Amendments: the Foundation for Experimental Evidence As the Basis for Drug Approvals

"substantial evidence" means evidence consisting of adequate and well controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the condition of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof.

1997 The Food and Drug Modernization Act (FDAMA); a Modification of the Substantial Evidence Criteria

It amends Section 505(d) of the Act by adding the following words: 'If the Secretary determines , based on relevant science, that the data from one adequate and well controlled clinical investigation and confirmatory evidence (obtained prior to or after such investigation) are sufficient to establish effectiveness, the Secretary may consider such data and evidence to constitute substantial evidence for purposes of the preceding sentence."

But what's wrong with one trial?

- Say, you had one very good trial with a type I error set at p = .05 and had planned for 80% power against negative results?

Theoretical Outcomes From a Series of 1000 Clinical Trials Assuming 1 of 10 Drugs Are Effective

100 trials with true efficacy: H_0 is false

$\beta = 0.20$

- **20 trials accept H_0**
- **80 trials reject H_0**

900 trials with no efficacy: H_0 is true

$\alpha = 0.05$

- **45 trials reject H_0**
- **855 trials accept H_0**

False positive trial rate = 45/(80 + 45) = 36%

Simon, R., Cancer Treatment Reports, V 66, 1982, 1083-1087

When and Why Would One Study Be Enough ?

- The Guidance for Industry discusses a number of issues of relevance
- What are the statistical implications ?
  - **multicenter studies**
  - **subgroups**
  - **consistency**
  - **statistically persuasive result**
    - **p-values, estimates of effect**

Evidence of Effectiveness from a Single Study (Guidance)

- Large multicenter study
  - **no single study/site provides unusually large fraction of patients**
  - **no single investigator or site provides a disproportionate favorable effect**
- Consistency across study subsets
- Multiple studies in a single study
- Multiple endpoints involving different events
- Statistically very persuasive finding

Evidence of Effectiveness from a Single Study (Guidance)

- Judgement
- Limited to situations where a clinically meaningful effect on:
  - **mortality**
  - **irreversible morbitity**
  - **prevention of a disease with a potentially serious outcome**

Case Study 1: “One trial”

- Tamoxifen for reductions in cancer incidence: the Breast Cancer Prevention Trial
  - One large (N=13,388) randomized controlled trial devoted to testing the hypothesis that tamoxifen reduces the incidence of breast cancer in asymptomatic but high risk women
  - Success of the trial led to approval: (note $p<.001$)
  - **ALONG WITH ...**

Case Study 1: “One trial”

- ***A massive amount of trial data available on secondary prevention of breast CA***
- Internally consistent results
- Extensive postmarketing information on adverse events
- No new toxicities detected

When Is One Study Sufficient ? Methodological Challenges

- What constitutes a very persuasive statistical finding?
- What is the likelihood the study reflects a true positive finding?
- What is the likelihood that, if repeated, a similar finding would be observed?
- Generalizability, extrapolation, replication - what do these mean?

Case Study 2: Bayesian Approaches to Device Trials: Choose your metaphor!

The Regulatory View

- Statutory directive for the FDA's CDRH:

  rely upon valid scientific evidence to determine whether there is reasonable assurance that the device is safe and effective.

- Valid scientific evidence is evidence from:
  - **well controlled studies**
  - **partially controlled studies**
  - **objective trials without matched controls**
  - **well documented case histories**
  - **reports of significant human experience (21 CFR 860.7)**

The 1997 Food and Drug Modernization Act (FDAMA): A Change in Approach to Device Studies?

"The Secretary shall consider, in consultation with the applicant, the least burdensome appropriate means of evaluating device effectiveness that would have a reasonable likelihood of resulting in approval."

Why Bayesian Medical Device Trials?

- One possible approach to realizing the least burdensome vision of Congress
- While each frequentist trial is *de novo*, a great deal of prior information on very similar devices often exists: clinical trials overseas, data registries, historical controls and pilots
- Use of good prior information can appreciably reduce the size and perhaps the length of a trial. One can arrive at the same decision in a much more timely manner.

TransScan T-200 Multi-frequency Impedance Breast Scanner

- Adjunct to mammography for women with BIRADS 3 or 4
- Approved April 16, 1999
- A single study based on the intended use on 72 women
- Strength borrowed from two other studies, a blinded study and a targeted study
- Bayesian multinomial logistic hierarchical model

TransScan T-200 Multi-frequency Impedance Breast Scanner

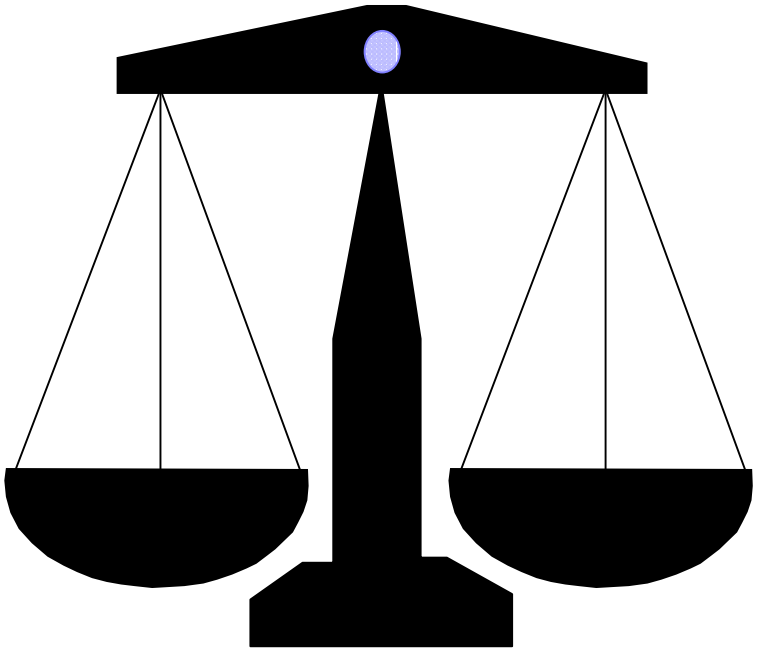
- Difference in sensitivity of T-2000 adjunct minus mammography 0.156 with a 95% credible interval (0.024, 0.288). Posterior prob. that diff. exceeds 0 is 0.99. Difference in specificity of T-200 adjunct minus mammography 0.202 with a 95% credible interval (0.009, 0.388). Posterior prob. that difference exceeds 0 is 0.98
- "Model predicts a substantial reduction in total number of negative biopsies, while increasing the net number of cancers detected."

Challenges Using Bayesian Inference in a Regulatory Setting

- Inference needs to be robust to range of models and priors that are reasonable to experts and consumers
- Need to insure a satisfactory replacement of type 1 error protection
- Compare performance of Bayesian and frequentist approaches to design and analysis
- Guard against presenting ‘best’ results among both classical and Bayesian alternatives
- Demonstrate reasonably consistent advantages of a Bayesian approach to the industry and the FDA
- More Bayesian clinical trial expertise is needed by companies and consultants

The Pre and Postmarket Balancing Act

- Another approach to realizing Congress's least burdensome vision may allow the use of postmarket data to speed premarket evaluation
- Postmarket problems and studies present many challenges



Case Study 3
Fetal Vacuum Extractors

- Approved as Class II pre-amendment
- Clinical indications:
  - Maternal
    - **Prolonged second stage of labor**
    - **Certain cardiac or cerebrovascular diseases**
    - **Inadequate expulsive efforts (certain pulmonary or neuromuscular disease)**
  - Fetal
    - **Nonreassuring fetal heart rate**

Premarket and Postmarket Device Related Data

- Data on randomized trials and other studies of safety and effectiveness
- Recent problems reported in the Medical Device Reporting (MDR) system and in the literature: severe adverse events and death

Data from NCHS on Assisted Vaginal Deliveries

Category	1991		1995	
Total deliveries	4,111,059		3,899,589	
Total deaths	35,496		29,505	
Total Forceps	183,116		134,718	
Forceps & death	(.36%)	664	(.35%)	472
Total Vacuum	176,392		228,364	
Vacuum & death	(.33%)	580	(.25%)	579

Regulatory Options

- Withdraw product from the market
- Release public health communication
- Carry out investigation using inspections
- Conduct research to address concerns

What are the Key Questions?

- Define relationship of indication to adverse outcomes
- Explore circumstances surrounding device use
- Fetal monitoring and evaluation of labor as influences in delivery method choice
- Find risk factors for major endpoints

Methodological Challenges?

- Rare exposure
- Rare adverse events
- Events can be very difficult to distinguish as related to vacuum extraction
- Complete data not usually available in medical records

Postmarket Paradoxes

- Efficacy very hard to study after market; safety a little easier to examine postmarket
- Pre to postmarket shift usually from tightly controlled to community environment
- Incentives and disincentives fall exactly contrary to the spirit of moving product to the market more swiftly
- Postmarket control expected by FDA but largely unrecognized

Summary

- **In all areas of medical product evaluation and postmarket monitoring, we find a host of methodological challenges**
- **In premarket evaluation, developing the minimally necessary but appropriate data to balance risks and benefits presents opportunities!**
- **How we can use existing data or postmarket monitoring to speed the premarket process while maintaining a high safety profile will challenge FDA, sponsors, CROs and the clinical community**